Convex recovery of tensors using nuclear norm penalization

Stéphane Chrétien [and] Tianwen Wei

Laboratoire de Mathématiques de Besancon
Université de Franche-Comté
16 route de Gray,
25000 Besancon, France

Abstract. The subdifferential of convex functions of the singular spectrum of real matrices has been widely studied in matrix analysis, optimization and automatic control theory. Convex analysis and optimization over spaces of tensors is now gaining much interest due to its potential applications to signal processing, statistics and engineering. The goal of this paper is to present an applications to the problem of low rank tensor recovery based on linear random measurement by extending the results of Tropp [6] to the tensors setting.

1 Introduction

1.1 Background

Tensors have been recently a subject of great interest in the applied mathematics community. We refer to [3,4] for a modern reference on this subject. Many applications of tensors are based on solving tensor related optimization problems, such as minimizing certain norms under linear constraints. Such problems have been recently successfully addressed in the 2D setting, i.e. for matrices, by the statistics, signal processing, inverse problems and automatic control communities in particular. Two of the reasons for this rapid growth of interest in the application of matrix norms to penalized estimation problems is that some norms promote spectral sparsity and that much work had been done in the fields of matrix analysis and convex analysis to analyze the subdifferential of such norms; see for example [7] and [5]. Our goal in the present paper is to extend previous results on matrix norms to the tensor setting. In particular, we propose a general study of the subdifferential of certain convex functions of the spectrum of real tensors and apply our results to the computation of the subdifferential of useful and natural matrix norms. We also present an application of our formulas to the problem of low rank tensor recovery using sparsity promoting norm minimization under random linear constraints, a natural extension of previous works by Tropp [6].

1.2 Notations

For any convex function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$, the conjugate function f^* associated to f is defined by

$$f^*(g) \stackrel{\text{def}}{=} \sup_{x \in \mathbb{R}^n} \quad \langle g, x \rangle - f(x).$$

The subdifferential of f at $x \in \mathbb{R}^n$ is defined by

$$\partial f \stackrel{\text{def}}{=} \{g \in \mathbb{R}^n \mid \forall y, \in \mathbb{R}^n \quad f(y) \geq f(x) + \langle g, y - x \rangle\}.$$

Moreover, it is well known (see e.g. [2]) that $g \in \partial f(x)$ if and only if

$$f(x) + f^*(g) = \langle g, x \rangle.$$

In the present paper, a tensor represented by a multi-dimensional array in $\mathbb{R}^{d_1 \times \cdots \times d_D}$. Let D and $n_1, \ldots, n_D$ be positive integers. Let $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ denote a D-dimensional tensor. If $n_1 = \cdots = n_D$, then we say that $\mathcal{X}$ is cubic. The set of D-mode cubic tensors will be denoted by $\mathbb{R}^{n \times \cdots \times n}$, where D will stay implicit. For any index set $C \subset \{1, \ldots, n_1\} \times \cdots \times \{1, \ldots, n_D\}$, $\mathcal{X}_C$ will denote the subarray $(\mathcal{X}_{i_1, \ldots, i_D})_{(i_1, \ldots, i_D) \in C}$.

2 Basics on tensors

2.1 Tensor norms

The spectrum of a tensor Let us define the spectrum as the mapping which to any tensor $\mathcal{X} \in \mathbb{R}^{n \times \cdots \times n}$ associates the vector $\sigma(\mathcal{X})$ given by

$$\sigma(\mathcal{X}) \stackrel{\text{def}}{=} \frac{1}{\sqrt{D}} \left(\sigma^{(1)}(\mathcal{X}), \ldots, \sigma^{(D)}(\mathcal{X}) \right),$$

where $\sigma^{(d)}(\mathcal{X})$ denotes the vector consisting of the singular values of the mode-d matricization of $\mathcal{X}$.

Norms of tensors Let $\mathcal{X} = (\mathcal{X}_{ijk})$ and $\mathcal{Y} = (\mathcal{Y}_{ijk})$ be tensors in $\mathbb{R}^{n_1 \times \cdots \times n_D}$. We can define several tensor norms on $\mathbb{R}^{n_1 \times \cdots \times n_D}$. The first one is a natural extension of the Frobenius norm or Hilbert-Schmidt norm from matrices to tensors. We start by defining the following scalar product on $\mathbb{R}^{n_1 \times \cdots \times n_D}$:

$$\langle \mathcal{X}, \mathcal{Y} \rangle \stackrel{\text{def}}{=} \sum_{i_1=1}^{n_1} \cdots \sum_{i_D=1}^{n_D} \mathcal{X}_{i_1,\ldots,i_D} \mathcal{Y}_{i_1,\ldots,i_D}.$$

Using this scalar product, we can define the following norm, which we call the Frobenius norm

$$\|\mathcal{X}\|_F \stackrel{\text{def}}{=} \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}.$$

One may also define an "operator norm" in the same manner as for matrices as follows

$$\|\mathcal{X}\| \stackrel{\text{def}}{=} \max_{\substack{u^{(d)} \in \mathbb{R}^{n_d}, \\ \|u^{(d)}\|_2=1, d=1,\ldots,D}} \langle \mathcal{X}, u^{(1)} \otimes \cdots \otimes u^{(D)} \rangle$$

We also define

$$\|\mathcal{X}\|_* \stackrel{\text{def}}{=} \frac{1}{D} \sum_{d=1}^{D} \|\sigma^{(d)}\|_1.$$

2.2 Orthogonally decomposable tensors

The Orthogonally decomposable (ODEC) tensors are defined as follows

Definition 2.1 *Let $\mathcal{X}$ be a tensor in $\mathbb{R}^{n_1 \times \cdots \times n_D}$. If*

$$\mathcal{X} = \sum_{i=1}^{r} \alpha_i \cdot u_i^{(1)} \otimes \cdots \otimes u_i^{(D)}, \tag{2.1}$$

where $r \leqslant n_1 \wedge \cdots \wedge n_D$, $\alpha_1 \geqslant \cdots \geqslant \alpha_r > 0$ and $\{u_1^{(d)}, \ldots, u_r^{(d)}\}$ is a family of orthonormal vectors for $d = 1, \ldots, D$, then we say (2.1) is an orthogonal decomposition of $\mathcal{X}$.

Denote $\alpha = (\alpha_1, \ldots, \alpha_r, 0, \ldots, 0)$ in $\mathbb{R}^{n_1 \wedge \cdots \wedge n_D}$. For each $d \in \{1, \ldots, D\}$, we may complete $\{u_1^{(d)}, \ldots, u_r^{(d)}\}$ with $\{u_{r+1}^{(d)}, \ldots, w_{n_d}^{(d)}\}$ so that matrix $U^{(d)} = (u_1^{(d)}, \ldots, u_{n_d}^{(d)}) \in \mathbb{R}^{n_d \times n_d}$ is orthogonal. Using $U^{(1)}, \ldots, U^{(D)}$, we may write (2.1) as

$$\mathcal{X} = \mathcal{D}(\alpha) \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_D U^{(D)}. \tag{2.2}$$

where $\mathcal{D} = \text{diag}(\alpha)$ is a diagonal tensor with the ith diagonal being α_i for $i = 1, \ldots, r$ and the other diagonal entries being zero. Note that representation (2.2) is generally not unique unless $n_1 = \cdots = n_D$ and $\alpha_1, \ldots, \alpha_r$ are all distinct.

It is easy to calculate the norms of ODEC tensors.

Proposition 2.2 *Let $\mathcal{X}$ be an orthogonally decomposable tensor and let*

$$\mathcal{X} = \sum_{i=1}^{r} \alpha_i \cdot u_i^{(1)} \otimes \cdots \otimes u_i^{(D)},$$

be an orthogonal decomposition of $\mathcal{X}$. Then

$$\|\mathcal{X}\| = \alpha_1 \quad \text{and} \quad \|\mathcal{X}\|_* = \sum_{i=1}^{r} \alpha_i.$$

3 Further results on the spectrum

In this section, we will present some further results on the spectrum such as the question of characterizing the image of the spectrum and the subdifferential of a function of the spectrum.

3.1 A technical prerequisite: Von Neumann's inequality for tensors

Von Neumann's inequality says that for any two matrices X and Y in $\mathbb{R}^{n_1 \times n_2}$, we have

$$\langle X, Y \rangle \le \langle \sigma(X), \sigma(Y) \rangle,$$

with equality when the singular vectors of X and Y are equal, up to permutations when the singular values have multiplicity greater than one. This result has proved useful for the study of the subdifferential of unitarily invariant convex functions of the spectrum in the matrix case in [5]. In order to study the subdifferential of the norms of certain type of tensors, we will need a generalization this result to higher orders. This was worked out in [1]. Let us recall the containt of the main result of [1].

Definition 3.1 *We say that a tensor $\mathcal{S}$ is blockwise decomposable if there exists an integer B and if, for all $d = 1, \ldots, D$, there exists a partition $I_1^{(d)} \cup \ldots \cup I_B^{(d)}$ into disjoint index subsets of $\{1, \ldots, n_d\}$, such that $\mathcal{X}_{i_1, \ldots, i_D} = 0$ if for all $b = 1, \ldots, B$, $(i_1, \ldots, i_D) \notin I_b^{(1)} \times \ldots \times I_b^{(D)}$.*

An illustration of this block decomposition can be found in Figure 1. The following result is a generalization of von Neumann's inequality from matrices to tensors. It is proved in [1].

Theorem 3.2 *Let $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ be tensors. Then for all $d = 1, \ldots, D$, we have*

$$\langle \mathcal{X}, \mathcal{Y} \rangle \leqslant \langle \sigma^{(d)}(\mathcal{X}), \sigma^{(d)}(\mathcal{Y}) \rangle. \tag{3.3}$$

Equality in (3.3) holds simultaneously for all $d = 1, \ldots, D$ if and only if there exist orthogonal matrices $W^{(d)} \in \mathbb{R}^{n_d \times n_d}$ for $d = 1, \ldots, D$ and tensors $\mathcal{D}(\mathcal{X}), \mathcal{D}(\mathcal{Y}) \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ such that

$$\mathcal{X} = \mathcal{D}(\mathcal{X}) \times_1 W^{(1)} \cdots \times_D W^{(D)},$$
$$\mathcal{Y} = \mathcal{D}(\mathcal{Y}) \times_1 W^{(1)} \cdots \times_D W^{(D)},$$

where $\mathcal{D}(\mathcal{X})$ and $\mathcal{D}(\mathcal{Y})$ satisfy the following properties:

(i) $\mathcal{D}(\mathcal{X})$ and $\mathcal{D}(\mathcal{Y})$ are block-wise decomposable with the same number of blocks, which we will denote by B,

(ii) the blocks $\{\mathcal{D}_b(\mathcal{X})\}_{b=1,\ldots,B}$ (resp. $\{\mathcal{D}_b(\mathcal{Y})\}_{b=1,\ldots,B}$) on the diagonal of $\mathcal{D}(\mathcal{X})$ (resp. $\mathcal{D}(\mathcal{Y})$) have the same sizes,

(iii) for each $b = 1, \ldots, B$ the two blocks $\mathcal{D}_b(\mathcal{X})$ and $\mathcal{D}_b(\mathcal{Y})$ are proportional.
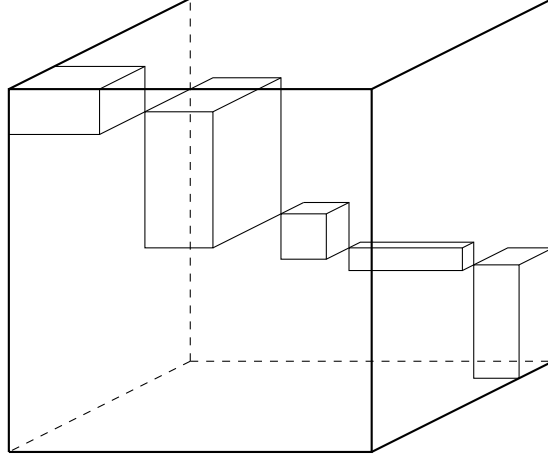
Fig. 1. A block-wise diagonal tensor.

3.2 Subdifferential for ODEC tensors

Theorem 3.3 *Let $f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \mapsto \mathbb{R}$ satisfy property*

$$f(s_1, \ldots, s_D) = f(s_{\tau(1)}, \ldots, s_{\tau(D)}) \tag{3.4}$$

for all $\tau \in \mathfrak{S}_S$. Then for all ODEC tensors $\mathcal{X}$, we have

$$(f \circ \sigma)^*(\mathcal{X}) = f^*(\sigma(\mathcal{X})) \tag{3.5}$$

Using this result combined with von Neumann's inequality for tensors, one easily obtains the following corollary.

Corollary 3.4 *Let $f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \mapsto \mathbb{R}$ satisfy property*

$$f(s_1, \ldots, s_D) = f(s_{\tau(1)}, \ldots, s_{\tau(D)}) \tag{3.6}$$

for all $\tau \in \mathfrak{S}_S$. Let $\mathcal{X}$ be an ODEC tensor. Then necessary and sufficient conditions for an ODEC tensor $\mathcal{Y}$ to belong to $\partial(f \circ \sigma)(\mathcal{X})$ are

1. $\mathcal{Y}$ has the same mode-d singular spaces as $\mathcal{X}$ for all $d = 1, \ldots, D$,
2. $\sigma(\mathcal{Y}) \in \partial f(\sigma(\mathcal{X}))$.

Corollary 3.5 *Let $\mathcal{X} = \mathcal{D}(\alpha) \times_1 U^{(1)} \times_2 \cdots \times_D U^{(D)}$ be an ODEC tensor. Then the subdifferential $\partial \| \cdot \|_*(\mathcal{X})$ includes the following set*

$$\Omega = \left\{ \mathcal{D}(\mathbf{1}) \times_1 U^{(1)} \times_2 \cdots \times_D U^{(D)} + \mathcal{V} \ \middle| \ \|\mathcal{V}\| \leqslant 1, \ \mathcal{V} \times_i U^{(i)\mathsf{T}} = 0, \ i = 1, \ldots, D \right\}.$$

4 Application to tensor recovery with gaussian measurements

Let $\mathcal{X}^\# \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an unknown true signal, $\Phi(\cdot) : \mathbb{R}^{n_1 \times n_2 \times n_3} \mapsto \mathbb{R}^m$ be a known linear measurement mapping and

$$y = \Phi(\mathcal{X}^\#) + \xi \tag{4.7}$$

be a noised vector of measurements in $\mathbb{R}^m$.

We focus on the following optimization problem:

$$\min_{\mathcal{X}} \|\mathcal{X}\|_* \quad \text{subject to } \|\Phi(\mathcal{X}) - y\| \leqslant \eta. \tag{4.8}$$

Let $\hat{\mathcal{X}}$ be any solution of optimization problem (4.8). We are interested in giving a bound for

$$\|\hat{\mathcal{X}} - \mathcal{X}^\#\|_F.$$

The main tool of this section is the following result by Tropp [6]:

Theorem 4.1 *Assume that $\|\xi\| \leqslant \eta$. Then with probability at least $1 - e^{-t^2/2}$, we have*

$$\|\hat{\mathcal{X}} - \mathcal{X}^\#\|_F \leqslant \frac{2\eta}{[\sqrt{m-1} - w(\mathscr{D}(\|\cdot\|_*, \mathcal{X}^\#)) - t]_+},$$

where $[a]_+ = \max\{a, 0\}$ for any $a \in \mathbb{R}$.

The quantity $w(\mathscr{D}(\|\cdot\|_*, \mathcal{X}^\#))$ denotes the conic Gaussian width $w(\cdot)$ of the descent cone $\mathscr{D}(\|\cdot\|_*, \mathcal{X}^\#)$. The definitions of these notions are given as follows:

Definition 4.2 *Let $K \in \mathbb{R}^d$ be a cone, the conic Gaussian width $w(K)$ is defined as*

$$w(K) = \mathbb{E}[\sup_{u \in K \cap \mathcal{S}^{d-1}} \langle g, u \rangle],$$

where $g \sim \mathcal{N}(0, I)$ is a standard Gaussian vector and $\mathcal{S}^{d-1}$ denotes the unit sphere in $\mathbb{R}^d$.

Definition 4.3 *Let $f : \mathbb{R}^d \mapsto \bar{\mathbb{R}}$ be a proper convex function. The descent cone $\mathscr{D}(f, x)$ of the function f at a point $x \in \mathbb{R}^d$ is defined as*

$$\mathscr{D}(f, x) \stackrel{\text{def}}{=} \{\lambda u \mid \lambda > 0, u \in \mathbb{R}^d, f(x + u) \leqslant f(x)\}.$$

According to Theorem 4.1, the error bound of $\|\hat{\mathcal{X}} - \mathcal{X}^\#\|_F$ depends on the conic Gaussian width $w(\cdot)$ of the descent cone $\mathscr{D}(\|\cdot\|_*, \mathcal{X}^\#)$. The following result reveals that the latter is then closely related to the subdifferential of $\|\cdot\|_*$ at $\mathcal{X}^\#$.

Proposition 4.4 *Assume that $\partial\|\mathcal{X}^\#\|$ is nonempty and does not contain the origin. Then*

$$w^2(\mathscr{D}(\|\cdot\|_*, \mathcal{X}^\#)) \leqslant \mathbb{E} \inf_{\tau \geqslant 0} \text{dist}_F^2(\mathcal{G}, \tau\partial\|\mathcal{X}^\#\|_*),$$

where $\mathcal{G} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a tensor with i.i.d. random Gaussian entries and

$$\text{dist}_F(\mathcal{G}, \tau\partial\|\mathcal{X}^\#\|_*) \stackrel{\text{def}}{=} \inf_{\mathcal{Y} \in \tau\partial\|\mathcal{X}^\#\|_*} \|\mathcal{G} - \mathcal{Y}\|_F,$$

i.e. the distance between $\mathcal{G}$ and the set $\tau\partial\|\mathcal{X}^\#\|_*$.

To derive a bound for $\|\hat{\mathcal{X}} - \mathcal{X}^\#\|_F$, we need to give an upper bound for

$$\mathbb{E} \inf_{\tau \geqslant 0} \text{dist}_F^2(\mathcal{G}, \tau\partial\|\mathcal{X}^\#\|_*).$$

The following result establishes such a bound in the case that $\mathcal{X}^\#$ is odec.

Proposition 4.5 *If $\mathcal{X}^\#$ is odec, then we have the following bound:*

$$\mathbb{E} \inf_{\tau \geqslant 0} \text{dist}_F^2(\mathcal{G}, \tau\partial\|\mathcal{X}^\#\|_*) \leqslant r^3 + r + 3r(n_1 + n_2 + n_3 - 3r) + r(n_1 n_2 + n_2 n_3 + n_1 n_3)$$

$$- r^2(n_1 + n_2 + n_3).$$

Proof. If $\mathcal{X}^{\#}$ is orthogonally decomposable, i.e.

$$\mathcal{X}^{\#} = \sum_{i=1}^{r} \sigma_i u_i^{(1)} \otimes u_i^{(2)} \otimes u_i^{(3)}$$
$$= \mathcal{D}(\sigma) \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)},$$

where $\mathcal{D}(\sigma)$ is a diagonal tensor with diagonal elements $\sigma = (\sigma_1, \ldots, \sigma_r)$ and $U^{(j)} = (u_1^{(j)}, \ldots, u_r^{(j)})$ for $j = 1, 2, 3$, then the subdifferential $\partial \| \cdot \|_*(\mathcal{X}^{\#})$ includes the following set

$$\Omega = \left\{ \sum_{i=1}^{r} u_i^{(1)} \otimes u_i^{(2)} \otimes u_i^{(3)} + \mathcal{V} \ \middle| \ \|\mathcal{V}\| \leqslant 1, \ \mathcal{V} \times_i U^{(i)} = 0, \ i = 1, 2, 3. \right\}. \tag{4.9}$$

Hence

$$\mathbb{E} \inf_{\tau \geqslant 0} \text{dist}_F^2(\mathcal{G}, \tau \partial \|\mathcal{X}^{\#}\|_*) \leqslant \mathbb{E} \inf_{\tau \geqslant 0} \text{dist}_F^2(\mathcal{G}, \tau \Omega) = \mathbb{E} \inf_{\tau \geqslant 0} \inf_{\mathcal{Y} \in \Omega} \|\mathcal{G} - \tau \mathcal{Y}\|_F^2.$$

Note that $\mathcal{V}$ in (4.9) can also be characterized by

$$\mathcal{V} = \mathcal{T} \times_1 U_{\perp}^{(1)} \times_2 U_{\perp}^{(2)} \times_3 U_{\perp}^{(3)}, \tag{4.10}$$

where $\mathcal{T} \in \mathbb{R}^{(n_1-r) \times (n_2-r) \times (n_3-r)}$ is a tensor such that $\|\mathcal{T}\| \leqslant 1$ and $U_{\perp}^{(i)} \in \mathbb{R}^{n_i \times (n_i-r)}$ is a matrix such that $\tilde{U}^{(i)} = (U^{(i)} | U_{\perp}^{(i)})$ is orthogonal for $i = 1, 2, 3$. In view of (4.9) and (4.10), we assert that any $\mathcal{Y} \in \Omega$ can be written as

$$\mathcal{Y} = \mathcal{C} \times_1 \tilde{U}^{(1)} \times_2 \tilde{U}^{(2)} \times_3 \tilde{U}^{(3)}.$$

where tensor $\mathcal{C}$ is block-wise diagonal with two diagonal blocks $\mathcal{C}_1 = \text{diag}(\mathbf{1}) \in \mathbb{R}^{r \times r \times r}$ and $\mathcal{C}_2 = \mathcal{T} \in \mathbb{R}^{(n_1-r) \times (n_2-r) \times (n_3-r)}$.

Because $\mathcal{G} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a tensor with i.i.d. random standard Gaussian entries, for any orthogonal matrices $W^{(1)}, W^{(2)}, W^{(3)}$ with appropriate size, tensor $\mathcal{G} \times_1 W^{(1)} \times_2 W^{(2)} \times_3 W^{(3)}$ still has i.i.d. standard Gaussian entries. Therefore, we may choose a coordinate system such that

$$\mathbb{E} \inf_{\tau \geqslant 0} \inf_{\mathcal{Y} \in \Omega} \|\mathcal{G} - \tau \mathcal{Y}\|_F^2 = \mathbb{E} \inf_{\tau \geqslant 0} \inf_{\mathcal{C} \in \tilde{\Omega}} \|\mathcal{G} - \tau \mathcal{C}\|_F^2,$$

where $\tilde{\Omega}$ denotes the set of block-wise diagonal tensors with two diagonal blocks $\mathcal{C}_{111} = \mathcal{D}(\mathbf{1}) \in \mathbb{R}^{r \times r \times r}$ and $\mathcal{C}_{222} \in \mathbb{R}^{(n_1-r) \times (n_2-r) \times (n_3-r)}$ verifying $\|\mathcal{C}_2\| \leqslant 1$. Partitioning $\mathcal{G}$ in the same manner, we obtain

$$\|\mathcal{G} - \tau \mathcal{C}\|_F^2 = \|\mathcal{G}_{111} - \tau \mathcal{D}(\mathbf{1})\|_F^2 + \|\mathcal{G}_{222} - \tau \mathcal{T}\|_F^2 + \sum_{\substack{i,j,k=1 \\ i,j,k \text{ are not equal}}}^{2} \|\mathcal{G}_{i,j,k}\|_F^2.$$

Since $\mathcal{G}$ is a tensor with independent Gaussian entries, it follows that

$$\mathbb{E} \sum_{\substack{i,j,k=1 \\ i,j,k \text{ are not equal}}}^{2} \|\mathcal{G}_{ijk}\|_F^2 = r(n_1 n_2 + n_2 n_3 + n_1 n_3) - r^2(n_1 + n_2 + n_3).$$

Thus

$$\mathbb{E} \inf_{\tau \geqslant 0} \inf_{\mathcal{C} \in \tilde{\Omega}} \|\mathcal{G} - \tau \mathcal{C}\|_F^2 = \mathbb{E} \inf_{\tau \geqslant 0} \inf_{\|\mathcal{C}_2\| \leqslant 1} \left(\|\mathcal{G}_{111} - \text{diag}(\tau)\|_F^2 + \|\mathcal{G}_{222} - \tau \mathcal{C}_2\|_F^2 \right)$$
$$+ r(n_1 n_2 + n_2 n_3 + n_1 n_3) - r^2(n_1 + n_2 + n_3).$$

Choosing $\tau = \|\mathcal{G}_2\|$, we get

$$\mathbb{E} \inf_{\tau \geqslant 0} \inf_{\|\mathcal{C}_2\| \leqslant 1} \left(\|\mathcal{G}_1 - \operatorname{diag}(\tau)\|_F^2 + \|\mathcal{G}_2 - \tau \mathcal{C}_2\|_F^2 \right) \leqslant \mathbb{E} \|\mathcal{G}_1 - \operatorname{diag}(\|\mathcal{G}_2\|)\|_F^2$$

Since

$$\mathbb{E}\|\mathcal{G}_1 - \operatorname{diag}(\|\mathcal{G}_2\|)\|_F^2 = r^3 + r\mathbb{E}\|\mathcal{G}_2\|^2 \leqslant r^3 + r + r\left(\sqrt{n_1 - r} + \sqrt{n_2 - r} + \sqrt{n_3 - r} \right)^2$$
$$\leqslant r^3 + r + 3r(n_1 + n_2 + n_3 - 3r),$$

It follows that

$$\mathbb{E} \inf_{\tau \geqslant 0} \operatorname{dist}_F^2(\mathcal{G}, \tau \partial \|\mathcal{X}^{\#}\|_*) \leqslant r^3 + r + 3r(n_1 + n_2 + n_3 - 3r) + r(n_1 n_2 + n_2 n_3 + n_1 n_3)$$
$$- r^2(n_1 + n_2 + n_3).$$

If the tensor is cubic, i.e. $n_i = n$ for $i = 1, 2, 3$, then we have with at least probability $1 - e^{-t^2/2}$ that

$$\|\hat{\mathcal{X}} - \mathcal{X}^{\#}\|_F \leqslant \frac{2\eta}{[\sqrt{m-1} - (r^3 + r + 9r(n-r) + 3rn(n-r)) - t]_+}.$$

References

1. Stéphane Chrétien and Tianwen Wei, *Von neumann's inequality for tensors*, arXiv preprint arXiv:1502.01616 (2015).
2. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms i: Part 1: Fundamentals*, vol. 305, Springer Science & Business Media, 1996.
3. Tamara G. Kolda and Brett W. Bader, *Tensor decompositions and applications*, SIAM Review **51** (2009), no. 3, 455–500.
4. J.M. Landsberg, *Tensors: geometry and applications*, vol. 128, American Mathematical Soc., 2012.
5. A. S. Lewis, *The convex analysis of unitarily invariant matrix functions*, Journal of Convex Analysis **2** (1995), no. 1, 173–183.
6. J. A. Tropp, *Convex recovery of a structured signal from independent random linear measurements*, arXiv preprint arXiv:1405.1102 (2014).
7. G Alistair Watson, *Characterization of the subdifferential of some matrix norms*, Linear Algebra and its Applications **170** (1992), 33–45.